DOCUMENT RESUME

ED 198 709 FL 012 090

AUTHOR Greenlee, Mel
TITLE Learning the Phonetic Cues to the Voiced-Voiceless Distinction: A Comparison of Child and Adult Speech Perception.
INSTITUTION Stanford Univ., Calif. Dept. of Linguistics.
PUB DATE Aug 78
NOTE 11p.: In its Papers and Reports on Child Language Development, Number 15, p160-169, Aug 1978. Not available in paper copy because of small type in original document.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adults: *Auditory Perception: *Child Language: Children: *Language Acquisition: Language Research: *Phonetics: Phonology: Psycholinguistics: Young Children
IDENTIFIERS *Voicing

ABSTRACT

Children and adults participated in two speech perception experiments in which they listened to recorded consonant-vowel-consonant (CVC) words in random order and responded by indicating which number of a word-pair they had heard. Some but not all stimuli that were intended to induce perceptual cross-over were successful in getting subjects to change the voicing category of their responses. Both vowel duration and subject age were found to significantly affect judgments about final obstruent voicing. Children showed perceptual cross-overs later in the duration continuum. These results are consistent with previous research. One result that remains to be explained, however, is the failure of the three-year-olds to shift their responses according to variations in vowel duration. The task may have been too difficult, or vowel duration differences in isolation may not have been sufficient cues to the final voicing contrast. The research is viewed as an inconclusive contribution to the issue of whether production or perception takes precedence. (Author/JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Stanford University Dept. of Linguistics

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Stanford Univ.

Dept. of Linguistics

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

160

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED198709

FL 012 090

Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception

Mel Greenlee
Phonology Lab
University of California-Berkeley

Introduction

Although a traditional belief in the study of child phonology has been that "perception precedes production" for phonological contrasts of the adult language, this generalization conceals a number of unanswered questions about early perceptual development. Studies of adult speech perception have shown that many phonetic or acoustic differences may signal one contrast at the phonological level. The purpose of research to be reported in this paper was to find out how children acquiring English learn to rank the phonetic differences in one instance of such a many-to-one situation. The phonological contrast at issue is the distinction between voiced and voiceless obstruents in word-final position. An example of this contrast is the minimal pair: back [bæk] vs bag [bæg].

This particular case of multiple cues seemed to be worth investigating for several reasons. First, there is a wealth of information available on adults' treatment or ranking of phonetic cues for [+ Voice] in this context. In spoken CVC pairs such as back-bag, although several acoustic differences can be associated with the final voicing contrast (see Greenlee 1977 for brief review), one of the most reliably produced differences is in vowel duration (Peterson and Lehiste 1960); other things being equal, the vowel in bag will be one-third to one-half again as long as the corresponding vowel in back.

In studies of adult speech perception, Denes (1955) and Raphael (1972) artificially created stimulus continua in which the final obstruents were uniformly voiceless, but in which vowel duration was carefully controlled. They found that, at least for these artificial stimuli, a sufficient increase in vowel duration would cause adult listeners to change or 'cross-over' from mostly voiceless to mostly voiced categorical judgments. From these results, it could be concluded that for adults, vowel duration is the most important, or primary, perceptual cue to [+ Voice] in final position [1].

The second reason concerns evidence from studies of children's early speech production (see e.g., Naeser 1970, Raphael, Dorman, and Geffner 1977, Smith 1976, Greenlee 1977). While neither children nor adults may produce many fully voiced final obstruents (Smith 1977), nevertheless children as young as age two reliably produce much longer vowels before obstruents which are pho-

nologically classified as [+ Voice]. From children's productions, it could be inferred that they also, like adults, rank vowel duration as the primary cue for differentiating [+ Voice] on word-final obstruents.

Finally, at the time I began the comparative research reported here, little information was available from direct studies of children's perception for the relevant acoustic differences---voicing and vowel duration [2]. Research on children's perception of specific acoustic cues continues to be relatively rare in comparison to studies of developing production. Koenigsknecht (1968) investigated children's ability to identify members of minimal pairs on the basis of vowel duration differences as part of a larger study of acoustic cue perception. Recently Zlatin and Koenigsknecht (1975) compared children and adults' ability to identify synthetic stimuli varied in voice-onset-time. The methodology of my research was very similar to that used by Zlatin and Koenigsknecht (as well as Krause and Fisher 1977), in that children of two ages and adults were asked to make forced-choice identifications of CVC words which were varied systematically in one acoustic parameter. In the present research, the parameter varied was vowel duration.

Method

Children and adults participated in two speech perception experiments in which they listened to tape-recorded CVC words presented in random order from a loudspeaker. Subjects responded by indicating which member of a word-pair they had heard. All sessions were conducted in a sound-treated room. As will be seen, Experiment I (Exp.I) was the main experiment, while Exp.II was on a much smaller scale.

Stimuli

Vowel durations of the experimental stimuli are given (in milliseconds) in Table 1.

Table 1
Vowel Duration of Test Stimuli

EXPERIMENT I

	Step 1	Step 2	Step 3	Step 4
BACK	171	204	274	344
BAG	204	159	124	88
BERT	150	194	259	318
BIRD	158	125	92	61

EXPERIMENT II

	Step 1	Step 2	Step 3	Step 4
BUS	154	187	254	321
BUZZ	199	160	127	94
CUP	78	129	191	253
CUB	112	83	54	23

All stimuli were originally tape-recorded natural utterances of an adult male English speaker. For the children, two sets of stimuli were used in both experiments: one set, represented in Table 1 above, will be referred to as the test set. Children also listened to a second, control set of words.

As can be seen in Table 1, eight CVC words were used in the two experiments; each of these eight original words was given four vowel durations in order to make the test stimuli. Test words were constructed by computer splicing, roughly as follows: (1) For originally voiced items (e.g., bag and bird in Exp.I), the voicing or voiced portion of the final consonant was first removed. Then vowel duration of the original word was cut back by successively splicing out pitch-periods from the vowel steady-state to create the four duration steps. (2) Originally voiceless items (e.g., back and Bert) were altered relatively less. To create duration steps from the voiceless words, pitch periods were repeated from the vowel steady-state to give four duration increments. It is important to keep in mind that none of the test stimuli contained voicing during the closure period. However, vowel durations included in the test stimuli were intended to encompass a range within which perceptual cross-over might be expected, at least from adult subjects. The corresponding control set of words for each of the experiments consisted of the eight (generally unaltered) CVC words as they had originally been recorded. Words in the control set thus contained both closure voicing and vowel duration differences as potential cues to the final voicing contrast, while test word 'steps' varied only in one dimension---vowel duration.

Table 2: Subjects

Experiment I

10 three-year-olds (mean age 2.11, range 2.1-3.3)
11 six-year-olds (mean age 6.6, range 5.2-6.9)
28 adults

Experiment II

2 three-year-olds (children listened only to
1 six-year-old variants of bus and buzz)
10 linguists
10 non-linguists

Subjects

Information about subjects is provided for each of the experiments in Table 2. The table shows that there were several differences between Exp.I and Exp.II, perhaps the most obvious being that Exp.I had more subjects. In Exp.I, age was the main difference between sub-

ject groups, but in Exp.II, adults were further separated into two groups on the basis of linguistic training (however, as it turned out, linguistic knowledge had no significant effect on adults' responses). In Exp.II, there were only three children, who listened only to variants for two of the four words.

Procedures

Children in each of the experiments participated in three types of listening sessions: training, test, and post-test sessions. In the training phase, children were introduced to a button-pushing game with a specially-constructed picture toy. In these initial sessions, children were required to differentiate maximal contrasts (e.g., back vs. bird) using the control set of words. When a word was heard from the loudspeaker, children were trained to respond by pushing the appropriate button, under one of two pictures on the toy, for which they received a reward. Figure 1 (see below) shows a child playing with the toy. The pictured words are back and bag.

After training on the maximally contrastive items, children went on to the 'test' sessions, in which they identified words varied only in vowel duration. For the test sessions, the two pictures on the toy represented a voiced-voiceless minimal pair.

Later, children identified the control set of words in a series of post-test sessions. The post-test was used to determine whether or not children could indeed differentiate between members of voiced-voiceless pairs in a more optimal situation, i.e. when both closure voicing and vowel duration differences were available as cues. Adults participated only in test sessions. Response mode for adult subjects was quite straightforward: they circled a word on an answer sheet, or wrote the corresponding final consonant. In the main experiment (Exp.I), adults and children identified 160 tokens of the test words (4 words, with 4 durations each, presented 10 times).

Results

The main aspect of subjects' identifications analyzed in statistical treatment of the data was the average number of 'voiced' responses each subject group gave for a particular duration variant, or 'step', of a word. Analyses of variance for effects of subject group and duration step were applied to the averaged data on voiced responses for each CVC word. When two-way ANOVA tests turned out to be significant, later one-way ANOVA and Scheffe post tests were used to probe for similarities among subject subgroups. While statistical analyses will not be described in detail, both effects of age and duration step were generally found to be significant in data from Exp.I. Main effects for duration step were signifi-

cant beyond the 0.001 level for back, bag, and Bert. Age effects were significant ($p<0.05$) for bag, Bert, and bird.

However, not all of the stimuli which were intended to induce perceptual cross-over were successful in getting subjects to change the voicing category of their responses. Ideally, when 'voiced' identifications are plotted as a function of increments in vowel duration, we should expect to see an S-shaped pattern (as in Denes 1955 or Raphael 1972).

Actual identification functions for the four words in each experiment are shown in Figures 2 and 3 below. Duration 'steps' for each word are arranged in such a way that vowel duration increases from left to right along the abscissa. Percentage of voiced identifications for each subject group is shown along the ordinate.

As can be seen in the figures, in each of the experiments, two of the test words were successful in eliciting perceptual cross-over, at least from adult subjects. In Exp. I, successful words were bag and Bert. In Exp. II, successful stimuli were those created from buzz and cup. Cub was overwhelmingly identified as voiceless 'cup', regardless of duration.

In addition to the differential success of stimulus words, Figures 2 and 3 illustrate important differences in the response patterns of subject groups. In Figure 2, it can be seen that adults and six-year-olds show a similar directional pattern for responses to the words bag and Bert: both subject groups cross over the 50% mark according to variations in vowel duration. On the other hand, the three-year-olds do not seem to change their responses very much for any word; responses for the youngest group tend to hover around the center (or chance area) of the graphs.

Differences in response pattern based on age were confirmed by one-way ANOVA tests. Adults and six-year-olds showed significant differences in the average number of 'voiced' identifications they gave according to duration step for the words bag and Bert ($p<0.05$), but for the three-year-olds, no significant differences according to duration were found for any word.

For those subject groups who did change the voicing category of their responses, a further comparison was carried out. For words which showed cross-over, vowel duration at the 50% mark (i.e., at the perceptual boundary between voicing categories) was determined by linear interpolation. These durations are shown in milliseconds in Table 3.

Figure 1: Child Playing with the Picture Toy

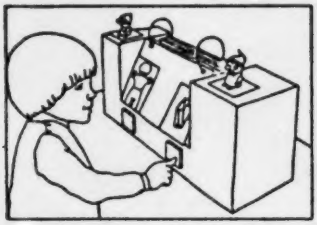

Figure 2
Exp. I: Voiced Identification as a Function of Duration

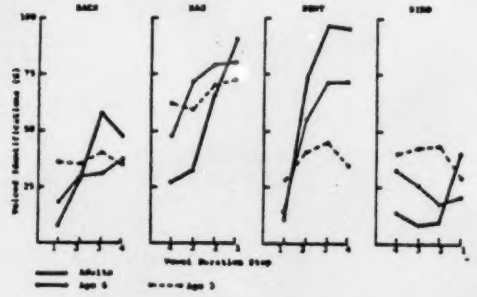

Figure 3
Exp.II: Voiced Identifications as a Function of Duration

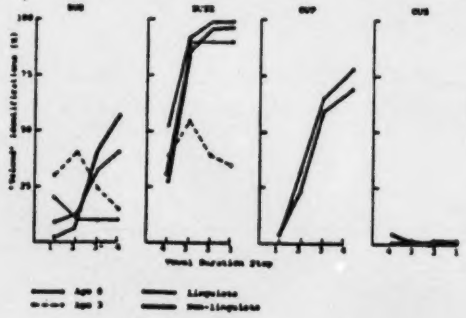

Table 3
Vowel Duration at the 50% Cross-over Point

	Adults	Six-year-olds
Voiceless to voiced BERT	177.5	189.4
Voiced to voiceless BAG	142.3	90.6

Table 3 shows that six-year-old subjects required relatively more modification of the original vowel duration in order to shift the voicing category of their responses; the cross-over point for children is later in the temporal continuum than that same point is for adults.

To summarize the outcome of the two experiments, both vowel duration and subject age were found to significantly affect judgments about final obstruent voicing. These results are not particularly surprising, based on findings of earlier comparative studies (e.g., Koenigsknecht 1968). The fact that children (in comparison to adults) showed perceptual cross-overs later in the duration continuum is consistent with recent research on voice-onset-time, in which the width of perceptual boundary was found to decrease with age (see Zlatin and Koenigsknecht 1975:549) Children may require a greater temporal change than do adults in order to shift their responses in discrimination tasks such as those used in the present study.

However, a puzzling result which remains to be explained is the failure of the three-year-olds to shift their responses according to variations in vowel duration. At least two explanations are possible: First, perhaps the task was too difficult for the youngest children, so they merely gave random responses. A second explanation would be based on acoustic cues in the stimulus items: For three-year-olds, vowel duration differences presented in isolation may not be sufficient cues to the final voicing contrast. Perhaps the younger children need more acoustic differences in order to make perceptual differentiations.

The first explanation (task difficulty) would predict that all stimuli, regardless of origin, would be identified randomly by the three-year-old subjects. The second explanation (based on acoustic parameters of the stimulus words) would imply that children should show more ability to differentiate [± Voice] if more acoustic cues, not just vowel duration, were available in the stimuli.

These predictions can be tested by comparing the accuracy of children's identifications for the test words, which varied only in vowel duration to their identifications for the corresponding control set of words, in which there were differences both in closure voicing and in vowel duration.

Figure 4
Children's Identifications for Test vs. Control Stimuli
(Percent correct responses)

Figure 4 shows the relative accuracy of children's identifications for the test and control words. For the originally voiced items, it can be seen that children's responses were more often accurate when closure voicing, as well as vowel duration, was present as a cue ($p<0.001$, in a binomial test of proportions). However, no significant differences between children's responses to the test versus control items were found for the originally voiceless words.

Results of the above comparison suggest that the explanation based on stimulus parameters is the more correct account of the three-year-olds' behavior. Both three-year-olds and six-year-old subjects apparently profit from the inclusion of more potential cues to [+ Voice], but the youngest children show this in the most extreme way. Three-year-olds seemed unable to differentiate [+ Voice] using only one acoustic cue, but their responses improved considerably when additional phonetic differences were included. In this regard, young children's responses are not without precedent, since similar identification patterns have been found with adult listeners. For example, Raphael (1972:1299) found that "stimuli synthesized with cues for voicing in the final segment were perceived as voiced following vowels of shorter duration than those synthesized with cues for voicelessness." Thus, differences between children's and adults' "identifications may reflect, not an absolute difference in whether or not multiple acoustic cues are used, but rather a difference in the extent to which they are used by the two age groups in discrimination tasks. However, a certain amount of caution is necessary in generalizing from children's responses in the data at hand, since it is always possible that tasks used in the identification sessions underestimated the three-year-olds' true perceptual abilities.

Children's performance on the identification tasks might appear surprising in light of the evidence (cited above) on early production of vowel duration differences. Although three-year-olds seemed unable to use vowel duration in isolation as a basis for perceptual identifications, vowel duration differences may in fact be the most reliably present acoustic cue to final voicing in young children's own early speech.

Conclusion

Despite these somewhat paradoxical results, the implications of this comparative study are not that, "production precedes perception," but that much more investigation of developing speech perception is needed. It seems likely that differences in response patterns of children and adults in the present study are based on developmental changes in the extent to which the complex of acoustic cues may be used in making perceptual discriminations. However, just as phonetic cues and phonological features need not show a one-to-one match, there need not be perfect parallelism between developing perceptual hierarchies and the acoustic differences evidenced in children's early speech production.

NOTES

[1] The contribution or perceptual weight of secondary acoustic differences has recently been investigated by Fruin and Bischoff (1976), using electronically spliced natural speech stimuli. They found that even when vowel duration was severely cut back (e.g., 72 ms), subjects could still identify a large percentage of intended final voiced stops as such, provided that other cues, such as closure voicing, were left intact. These results indicate that adults' perceptual hierarchies may be somewhat flexible; i.e., adults may rearrange the ranking of phonetic cues in order to fit the requirements of particular speech perception tasks.

[2] After data collection for the present study had been completed, I learned of additional relevant research on early perception, which can only be summarized briefly in this report. Krause and Fisher (1977) conducted a comparative experiment which was very similar to the present investigation. However,in their research, vowel duration variants were created through speech synthesis (as in Zlatin and Koenigsknecht 1975), and duration represented a larger range than CVC test words listed in Table 1. Krause and Fisher obtained perceptual cross-overs from all subject groups (age 3, age 6, and adults), although effects of age on cross-over point were similar to those illustrated in Table 3.

Higgs and Hodson (1978) found that four-year-old children did not identify members of voiced-voiceless na-

tural speech pairs as well as as adult listeners. In particular, children often made errors in 'decoding' voiceless final obstruents, tending to confuse them with their voiced cognates.

REFERENCES

Denes, P. 1955. Effect of duration on the perception of voicing. Journal of the Acoustical Society of America (JASA) 27.761-764.

Fruin, C. & Bischoff, D. 1976. Vowel duration as a cue for consonant voicing? Presented to the Acoustical Soc. of America, San Diego, Calif.

Greenlee, M. 1977. Learning the phonetic cues to the voiced-voiceless distinction: Preliminary study of four American English speaking children. Presented to the Calif. Linguistics Assoc. Conference, Fresno.

Higgs, J.W. & Hodson, B. 1978. Phonological perception of word-final obstruent consonants. J. Phonetics 6.25-35.

Koenigsknecht, R.A. 1968. An investigation of the discrimination of certain spectral and temporal acoustic cues for speech sounds in three-year-old children, six-year-old children, and adults. PhD diss., Northwestern U.

Krause, S. & Fischer, H. 1977. Developmental use of preceding vowel duration as a cue to postvocalic consonant voicing: A perception study. Presented to Amer. Speech and Hearing Assoc., Chicago.

Naeser, M. A. 1970. The American child's acquisition of differential vowel duration. U. Wisconsin Res. & Dev. Center, Tech. Report 144.

Peterson, G.E. and Lehiste, I. 1960. Duration of syllabic nuclei in English. JASA 51. 1296-1303.

Raphael, L.J. 1972. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. JASA 51.1296-1303.

Raphael, L.J., Dorman, M. & Geffner, D. 1977. Voicing-conditioned durational differences in vowels and consonants in the speech of three-and four-year-old children. Haskins Status Report on Speech Res. 50.115-122.

Smith, B.L. 1976. Developmental aspects of some temporal parameters of English speech production. PhD diss., U. Texas.

Smith, B.L. 1977. Phonetic aspects of "devoiced" stops in children's speech. Presented to the Acoustical Soc. Am., Miami.

Zlatin, M.A. & Koenigsknecht, R.A. 1975. Development of the voicing distinction: Perception of stop consonants. JSHR 18.541-553.